5. *Skewness*

Meaning and Definition

Literal meaning of skewness is **'lack of symmetry'**. It is a numerical measure which reveals asymmetry of a statistical series.

> *According to **Paden** and **Lindquist**, "A distribution is said to be skewed if it is lacking in symmetry, that is, if the measure tend to pile up at one end or the other, of the range to measures."*

> *In the words of **G. Simpson** and **F. Kafka**, "Skewness or asymmetry is the attribute of frequency distribution that extends further on one side of the class with the highest frequency on the other."*

> ***Morris Humburg** said, "Skewness refers to the asymmetry or lack of symmetry in the shape of a frequency distribution. This characteristic is of particular importance in connection with judging the typicality of certain measures of central tendency."*

> *Similarly **Croxton** and **Cowden** defined it, "When a series is not symmetrical, it is said to be asymmetrical or skewed."*

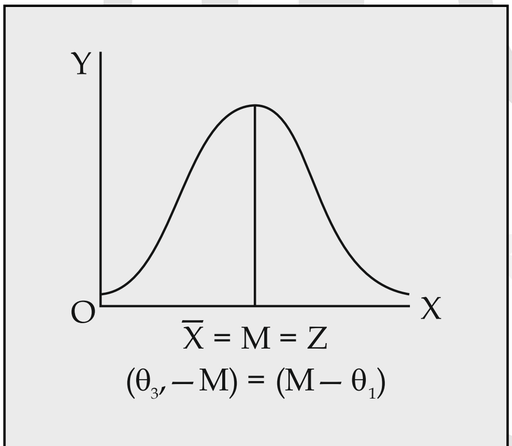

Fig. (a) : *Normal or Symmetrical*

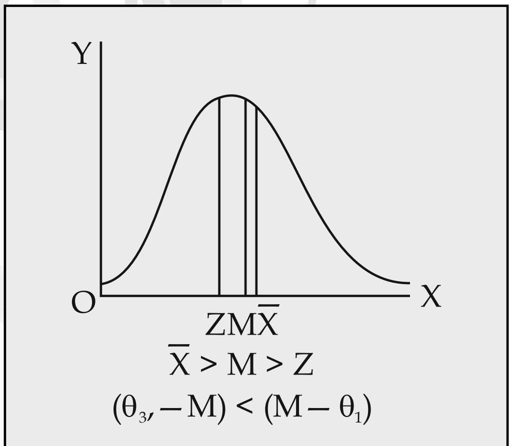

Fig. (b) : *Positive Skewness*

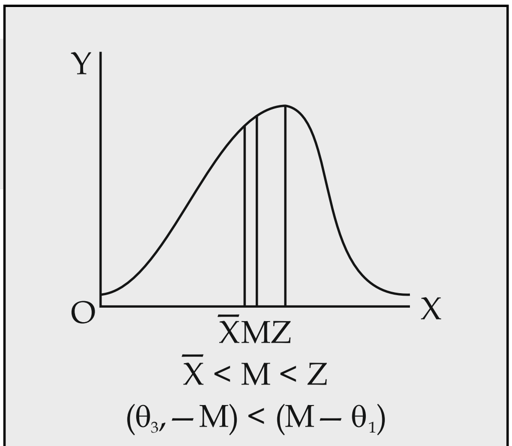

Fig. (c) : *Negative Skewness*

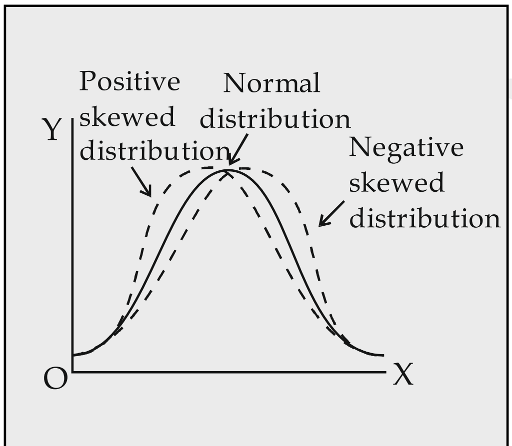

Fig. (d) : *Skewness*

Fig. (b) reveals the shape of moderately skewed curve. It is tilted towards right. In this case value of mean would be more than the values of median and mode. Median value would be higher than the value of mode. This curve depicts positive skewness.

Fig. (c) also reveals moderately skewed distribution. It is tilted towards left. In this case value of mode will be greater than the value of median, and the value of median will be greater than the value of mean. Thus such curve depicts negative skewness.

On the basis of first three figures a combined figure (d) can be drawn as under to reveal all three situations :

Thus

(I) In a symmetrical distribution $\overline{X} = M = Z$

(ii) In a positively skewed distribution $\overline{X} > M > Z$

(iii) In a negatively skewed distribution $Z > M > X$

In different types of frequency distribution the position of measures of central tendency may be presented as under :

Different Types of Skewness

Size	A Frequency	B Frequency	C Frequency
5	100	100	100
10	300	900	200
15	500	500	300
20	700	400	400
25	500	300	800
30	300	200	900
35	100	100	100
Skewness	No Skewness	Positive Skewness	Negative Skewness
Position of Averages	$\overline{X} = M = Z(20)$	$\overline{X} > M > Z$ $16.8 > 15 > 10$	$\overline{X} < M < Z$ $23.2 < 25 < 30$
Quartiles	$(Q_3 - M) = (M - Q_1)$	$(Q_3 - M) > (M - Q_1)$	$(Q_3 - M) < (M - Q_1)$
Curve	Normal	Skewed towards right	Skewed to the left

Test of Skewness : With an object to ascertain whether the distribution is normal or skewed the following facts should be considered :

(1) Relationship between Averages : If in a distribution the values of Mean, Median and Mode are not identical, it is a skewed distribution. The greater the difference between Mean and Mode, distribution will be more skewed.

(2) Distance of Pair of Quartiles from Median : If in a distribution the values of Q_3 and Q_1 are equi-distant from median value, it is a symmetrical distribution. If they are not equi-distant, it is a skewed distribution.

(3) Frequencies on either sides of Mode : If the total frequencies on both sides of the modal value are not equal, it is a skewed distribution.

(4) **Total of Deviations :** If the sum of positive deviations from the value of Median or Mode are equal to the sum of negative deviations, there is no skewness in the distribution.

(5) **The Curve :** When the data of a distribution are plotted on a graph paper and if the curve is not bell-shaped (normal), it is a skewed distribution.

Difference between Dispersion and Skewness

Basis	Dispersion	Skewness
1. Nature	These measures depict the scatteredness or spread of values from a measure of central tendency.	Measures of skewness show whether the series is symmetrical or asymmetrical. It indicates the shape of the frequency curve.
2. Base	Measures of dispersion depend upon the averages of the second order.	Measures of skewness depend upon the averages of first and second order.
3. Variability	Measures of dispersion do not reveal on which side of the central value the values have more variations.	Skewness indicate the direction of variation, i.e., variations on the two sides of median are known.
4. Conclusion	All the measures of dispersion are positive	Coefficient of skewness can be positive or negative.
5. Relation with Moments	Measures of dispersion are based upon all the three moments of mean.	Measures of skewness are based on first and third moments only.
6. Normal Distribution	A normal distribution may have some value of dispersion.	Measure of skewness is zero, i.e, there is no skewness in normal distribution.
7. Presentation	Dispersion cannot be presented by means of diagrams.	Skewness can easily be presented by diagrams.

Measures of Skewness

Measures of skewness are the devices to find out the direction and the extent of asymmetry in a statistical series. In other words, measures of skewness are meant to give an idea about the direction and degree of asymmetry in a variable. These measures can be absolute or relative. Absolute measures of skewness tell us the degree of asymmetry and whether it is positive or negative in value. This measure is not suitable for comparative study. As such to compare the degree of skewness between two or more distributions relative measures of skewness are computed.

It means, absolute measure of skewness is changed into relative measure for the purpose of comparison. The relative measure is called 'Coefficient of Skewness' which is represented by '**J**'.

There are four measures of skewness :

1. *Karl Pearson's measure*
2. *Bowley's measure*
3. *Kelly's measure*
4. *Moments measure*

(1) Karl Pearson's Measure

This measure is based on Mean and Mode. When Mode is ill defined in a distribution then Median is used in place of Mode. To compute the relative measure or coefficient, absolute measure is divided by standard deviation. Symbolically,

(i) Absolute Measure :

$$\boxed{\text{Skewness (Sk)} = \text{Mean}(\bar{X}) - \text{Mode}(Z)}$$

$$\boxed{\text{Coefficient of Skewness (J)} = \frac{\text{Mean}(\bar{X}) - \text{Mode}(Z)}{\text{S.D.}(\sigma)}}$$

If mode is ill-defined,

$$\boxed{\text{Skewness (Sk)} = 3[\text{Mean}(\bar{X}) - \text{Median}(M)]}$$

$$\boxed{\text{Coefficient of Skewness (J)} = \frac{3(\text{Mean} - \text{Median})}{\text{S.D.}(\sigma)}}$$

The second formula is based on empirical relationship between averages.

(ii) Relative Measure or Coefficient of Skewness (J) :

$$\boxed{\text{(a) } J = \frac{\bar{X} - Z}{\sigma} \quad \text{(b) } J = \frac{(\bar{X} - M)}{\sigma}}$$

where J = Coefficient of skewness

$\bar{X}$ = Mean

M = Median

Z = Mode

σ = Standard deviation

Example : The following information is related to two distributions, state which distribution is more skewed :

	Distribution I	Distribution II
Mean	100	90
Median	95	95
Standard Deviation	10	10

Solution : The value of Mode is not given in this problem, and so it will be solved by 2nd formula given by Karl Pearson.

Distribution I : $J = \frac{3(\overline{X} - M)}{\sigma} = \frac{3(100 - 95)}{10} = +1.5$

Distribution II : $J = \frac{3(\overline{X} - M)}{\sigma} = \frac{3(90 - 95)}{10} = -1.5$

Thus both the distributions reveal the same amount of skewness. But it is positive in distribution I while negative in distribution II.

Example : From the information given below, find out which group is more skewed.

Group X	100	118	122	109	105	107	121	113	105	100
Group Y	18	26	21	25	23	22	17	15	16	17

Solution : It is a case of individual series where mode is not certain, and so we will use median instead of mode.

Calculation of Skewness by $\frac{3(\overline{X} - M)}{\sigma}$ **formula :**

S.No.	**Group X**			**Group Y**		
	Ascending Array	**Deviation from $\overline{X} = 110$**	**Square of Deviation**	**Ascending Array**	**Deviation from $\overline{Y} = 20$**	**Square of Deviation**
	X	dX'	d^2X	Y	dY'	d^2Y
1	100	– 10	100	15	– 5	25
2	100	– 10	100	16	– 4	16
3	105	– 5	25	17	– 3	9
4	105	– 5	25	17	– 3	9
5	107	– 3	9	18	– 2	4
6	109	– 1	1	21	+ 1	1
7	113	+ 3	9	22	+ 2	4
8	118	+ 8	64	23	+ 3	9
9	121	+ 11	121	25	+ 5	25
10	122	+ 12	144	26	+ 6	36
N = 10	$\Sigma X = 1100$		$\Sigma d^2X = 598$	$\Sigma Y = 200$		$\Sigma d^2Y = 138$

Group X	Group Y
$\bar{X}_x = \frac{\Sigma X}{N} = \frac{1100}{10} = 110$	$\bar{X}_y = \frac{\Sigma X}{N} = \frac{200}{10} = 20$
$\sigma_x = \sqrt{\left(\frac{\Sigma dX}{N}\right)^2} = \sqrt{\left(\frac{598}{10}\right)} = 7.733$	$\sigma_y = \sqrt{\left(\frac{\Sigma dY}{N}\right)^2} = \sqrt{\left(\frac{138}{10}\right)} = 3.715$
M_x = size of $\frac{N+1}{2}$ or $\frac{11}{2}$ or 5.5^{th} item	M_y = size of 5.5^{th} item
$= \frac{107+109}{2} = 108$	$= \frac{18+21}{2} = 19.5$
$J_x = \frac{3(\bar{X}-M)}{\sigma} = \frac{3(110-108)}{7.733}$	$J_y = \frac{3(\bar{X}-M)}{\sigma} = \frac{3(20-19.5)}{3.715}$
$= \frac{6}{7.733} = +0.776$	$= \frac{1.5}{3.715} = 0.404$

Skewness is positive in both the groups, but group x is more skewed.

(2) Bowley's Measure

Dr. A.L. Bowley propounded another measure of skewness based on the relative position of median and the two quartiles. If a distribution is symmetrical then Q_1 and Q_3 will be at equal distance from the Median. If a distribution is asymmetrical, the quartiles will not be equi-distant from the Median. Larger the difference, higher would be the degree of skewness. Bowley's measure of skewness is called second Measure of Skewness or Quartile Measure of Skewness. This measure is useful in distributions where mode is ill-defined. This measure can also be used in open end distributions. Its formula is as under :

Bowley's Measure of Skewness or Quartile Measure :

$$\boxed{Sk = (Q_3 - M) - (M - Q_1) \text{ or } Q_3 + Q_1 - 2M}$$

Bowley's Measure of Coefficient of Skewness or Quartile Measure :

$$\boxed{J_Q = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} \text{ or } \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}}$$

Example : Find out quartile coefficient of skewness from the following distribution :

Height (cm.)	75	76	77	78	79	80	81	82	83
No. of Students (f)	6	8	13	18	20	16	10	7	2

Solution :

Calculation of Bowley's Coefficient of Skewness

Heght (cm.) (X)	No. of Students (f)	Cum. Freq. (cf)
75	6	6
76	8	14
77	13	27
78	18	45
79	20	65
80	16	81
81	10	91
82	07	98
83	02	100
Total	100	

M = Value of $\left(\frac{N+1}{2}\right)^{th}$ item or $\frac{101}{2}$ = 50.5^{th} item. It lies in 65 c.f., Value = 79

Q_1 = Value of $\left(\frac{N+1}{4}\right)^{th}$ item or $\frac{101}{4}$ = 25.25^{th} item. It lies in 27 c.f. Value = 77

Q_3 = Value of $\left(\frac{3N+1}{4}\right)^{th}$ item or $\frac{3\times101}{4}$ = 75.75^{th} item. It lies in 81 c.f., value = 80.

$$J_a = \frac{Q_3+Q_1-2M}{Q_3-Q_1}$$

$$= \frac{80+77-(2\times79)}{80-77}$$

$$= \frac{157-158}{3}$$

$$= -0.333$$

Skewness is negative.

(3) Kelly's Measure

Kelly's measure of skewness is based on middle 90% of observations as against Bowley's measure where middle 50% observations are taken into account. It is a mid-way between Karl Pearson's measure and Dr. Bowley's measure. According to him :

Focus Formula

$$\text{Skewness } Sk = P_{90}+P_{10}-2P_{50} \text{ or } D_9+D_1-2D_5$$

$$\text{Coeff. of } Sk\,(J_p) = \frac{P_{90}+P_{10}-2P_{50}}{P_{90}-P_{10}} \text{ or } \frac{D_9+D_1-2D_5}{D_9-D_1}$$

This measure is rarely used in actual practice.

Example : From the following data given for a frequency distribution, calculate Kelly's Coefficient of Skewness :

$P_{90} = 101,\ P_{10} = 58.12,\ P_{50} = 79.06$

Solution :

$$J_P = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}} = \frac{101 + 58.12 - 2(79.06)}{101 - 58.12} = \frac{159.12 - 158.12}{42.88} = +0.023$$

(4) Moments Measure of Skewness

This measure of skewness is based on moments about dispersion. Prof. King treated this measure as the best but its computation is very hard and complicated. It is because of its complexity, that its use is restricted. It is very rarely used in actual practice. It is obtained with the help of squares cubes etc. of the values. This measure is based on the assumption that the sum of deviation of values from the mean is zero. Its absolute measure is the cube-root of third moment of dispersion.

Ques. Which one of the following is a false description ? **(NTA UGC-NET July 2016 P-II)**

(1) In a moderately asymmetrical distribution, the empirical relationship between Mean, Mode and Median suggested by Karl Pearson is
Mean – Mode = 3 (Mean – Median)

(2) Coefficient of variation is an absolute measure of dispersion

(3) Measure of skewness indicates the direction and extent of skewness in the distribution of numerical values in the data set

(4) Kurtosis refers to the degree of flatness or peakedness in the region around the mode of a frequency curve

Ans. (2) Coefficient of variation is an absolute measure of dispersion.